Linear functions

- Graphically, they are represented by straight lines.
- The $x$-intercept, or horizontal intercept, is the $x$-coordinate of the point where the graph meets the horizontal axis; the input for which the output is 0.
- The $y$-intercept, or vertical intercept, is the $y$-coordinate of the point where the graph meets the vertical axis; the output for an input of 0.
- The slope of the line is a measure of its inclination or steepness:

$$
\text{slope} = \frac{\text{rise}}{\text{run}}
$$

graphs of increasing linear functions have positive slope, those of decreasing linear functions have negative slope, and constant linear functions have 0 slope.

- Slope can be measured between any two points on a line; but since

$$
\text{rise} = \text{change in the output values} \\
\text{run} = \text{change in the input values}
$$

it also measures the average rate of change of the function (measured in output units per input unit) over that interval of inputs.
The slope is the same between every pair of points on the line (why?), so the rate of change of a linear function is constant over all input intervals

• Numerically, linear functions are recognizable by the property that slope computations are the same between any pair of input/output values
• In the common case in which tabulated values involve equally spaced inputs, this property is equivalent to requiring that outputs are also equally spaced, that is, that they have equal \textbf{first differences}
• For real data, perfect linearity is rare to find: data with equally spaced inputs and \textit{nearly} equal first differences are well-modeled by a linear function
• Many real data sets involve annual measures of some quantity: the input is time, measured in years; because year numbers are large quantities (near 2000), it is best not to use these numbers to measure the time variable, but to \textbf{align} the data by choosing a convenient year as year 0 and measuring relative to this year

• Symbolically, linear functions are characterized by the form of a linear equation: \( y = f(x) = ax + b \)
• The \textbf{parameters} \( a \) and \( b \) have clear interpretations:
  \( a \) is the slope and rate of change of the function
  \( b \) is the \( y \)-intercept and output for 0 input
The linear regression model

Given data that is not perfectly linear, we can find the linear function that best “fits” the data by a process called linear regression, or the method of least squares:

- suppose \((x_{\text{data}}, y_{\text{data}})\) is one of the data points and \(f(x)\) is a linear function
- then \(y_{\text{line}} = f(x_{\text{data}})\) is equal to \(y_{\text{data}}\) only if the line passes through the point
- if the function is meant to model the data, then the difference \(y_{\text{data}} - y_{\text{line}}\) is the error made by the model at that data point
- the function that best fits the data minimizes the overall error with the entire set of data values
- we quantify this overall error as the sum of the squared errors (SSE) between the line and each of the data points (there are other ways to do this, but this is the most commonly used method)

[TI-83: STAT  CALC
LinReg(ax+b) <x-list>, <y-list>, <Y-variable>
produces the linear regression model]
Numerical accuracy in models

• Your calculator generally computes up to 10-digit accuracy, but this does not necessarily mean that all these digits are accurate: *the results of computations are no more accurate than the least accurate values used to generate the results.*

• When you record numbers associated with a model or with computations from a model, *round to an appropriate number of digits of accuracy* (typically, four digits is more than sufficient).

• However, *never round values until the necessary computations have been completed*, otherwise the rounding error will affect the computational accuracy.

• When reporting computations associated with a model, *appropriate units must accompany the numbers to give them meaning.*

• When recording a model formula, *indicate what interval of input values are associated with the model*; using the model for inputs within this interval is called *interpolation*, while using the model for inputs outside the interval is *extrapolation* (the further outside the interval, the more unreliable are such calculations).